脉冲双星加速度搜索方法及软件改进

冀昶旭，刘志杰，于徐红
贵州省信息与计算科学重点实验室

摘要：毫秒脉冲星对引力波探测、太空导航等方面具有重要意义。本文通过引证 ATNF 数据库，指出双星系统与毫秒脉冲星存在密切关系。分析了 FAST 早期科学数据中心脉冲双星搜索能力的现状. 目前 Ransom 版本的 accelsearch 算法虽在 CPU 上进行了高度优化，但在 FAST 庞大数据量下，仍十分耗费算力。罗晋涛 GPU 版本的加速度搜索带来 6-8 倍的加速效果，但未能针对 GPU 内存进行有效优化。本文先概述脉冲双星及其轨道运动对信号处理带来的影响，然后引出相应的搜索技术及原理，为提高脉冲双星的搜索效率，利用并行化技术编写程序对 GPU 版本的加速度搜索进行优化。结果表明，使用不同 zmax 参数对不同.dat 文件进行处理，速度均有明显提升。

关键词：加速度搜索；脉冲双星；FAST

中图分类号：P162
基金项目：国家自然科学基金（U1631132，U1731238，U1831131）

作者简介：
冀昶旭，男，硕士研究生，研究方向：FAST 数据处理，Email: 18010210502@gznu.edu.cn
刘志杰，男，教授，研究方向：FAST 数据处理，Email: liuzj@gznu.edu.cn
于徐红，男，高级工程师，研究方向：FAST 数据处理，Email：yuxuhong@gznu.edu.cn

脉冲双星系统的发现极具科学意义。1974 年，赫尔斯和泰勒在天鹰座天域探测到脉冲信号，经计算发现是一颗脉冲双星[1]（PSR B1913+16）的信号。两人通过深入研究首次发现引力波存在的间接定量证据[2]，并因此获得 1993 年诺贝尔物理学奖。2020 年 2 月，王琳等基于 FAST 数据在 M13 中发现一颗处于双星系统中的毫秒脉冲星 M13F，并认证该星团中另一脉冲星 M13E 为掩食双星，首次测量出 M13 中 4 个脉冲双星系统的轨道椭率，同时获得 M13 现有脉冲星的国际最好计时结果[3]。这些成果都在推动着科学的进步。

遗憾的是，现有脉冲双星的搜索较为耗时，搜索能力亟待提升。一方面是 FAST 早期科学数据中心对数据存储和处理能力的要求。在 FAST 开启 19 波束巡天漂移扫描后，数据量激增，每天产生压缩数据 50TB，每年（按 200 天计算）将产生 10PB 超级数据，如何利用好这批数据，是对数据中心的存储和超算能力的严峻考验。另一方面，高效的搜索能力对 FAST 探测优质的毫秒级脉冲星具有重要意义。在 FAST 早期科学数据中心①脉冲星分布式搜索计算系统 Craber[4]中使用加速度搜索（accelsearch）方法[5-7]，当参数 zmax 设置较大时（一般为 200 以上），搜索效率显著降低。目前，Ransom 版本的 accelsearch 方法[8]虽然在 CPU 上进行了高度优化，但速度仍有所欠缺。单机八核处理一个大小约 140G 的巡天漂移扫描数据，zmax 参数预设 200，耗时约～20h，这仅是 10min 巡天且经过处理的数据量。显然，这种数据处理能力极大限制了脉冲双星的有效发现。

此外，双星系统中蕴含大量毫秒级脉冲星。如表 1 和图 1 中数据，球状星团中包含大量脉冲双星，且多为毫秒级脉冲星。截止 2020 年底，ATNF 脉冲星数据库共收录脉冲双星 149 颗②。其中约有 130 颗在球状星团中发现，表 1 中给出部分球状星团中样本数据统计结果，图 1 统计了 148 颗脉冲星的周期分布（有 1 颗周期为 2.7s 统计意义不大未记入）。

表 1 截止 2020 年 12 月部分球状星团的样本数据②

Table 1 Sample data of part globular clusters till December 2020

Cluster Name	Galactic longitude (deg)	Galactic latitude (deg)	Distance (kpc)	Number of Pulsars	Number of binary pulsars
47 Tuc (NGC 104)	305.90	−44.89	4.5	25	15
M3 (NGC 5272)	42.41	78.71	10.4	> 3	> 2
M5 (NGC 5904)	3.86	46.80	7.5	5	4
M13 (NGC 6205)	59.01	40.91	7.7	5	3
M62 (NGC 6266)	353.57	7.32	6.9	6	6

① 2014 年，国家天文台与贵州师范大学共同建立 FAST 早期科学数据中心
② http://www.atnf.csiro.au/people/pulsar/psrcat/

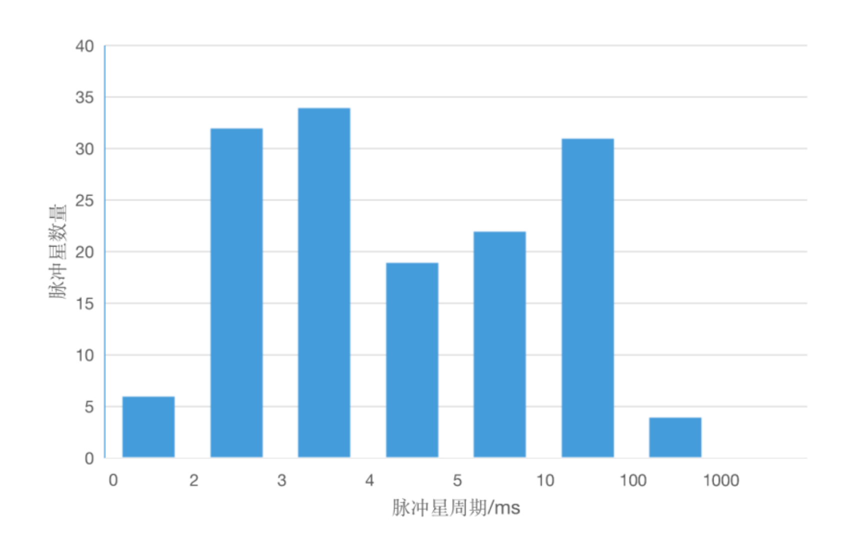
Terzan 5	3.84	1.69	10.3	38	19
NGC 6440	7.73	3.80	8.4	6	3
NGC 6441	353.53	−5.01	11.7	4	2
NGC 6544	5.84	−2.20	2.7	2	2
M28 (NGC 6626)	7.80	−5.58	5.6	12	8

在距太阳系～15kpc 内球状星团的脉冲星中，双星数量占比较大。其中，47 Tuc 形状球团已发现的脉冲星中，60%为双星。Terzan 5 经认证的脉冲星中有一半为脉冲双星。

Among the pulsars of globular clusters within ~15kpc of the solar system, the number of binary pulsar is relatively large. Among the pulsars discovered in 47 Tuc, 60% are binary pulsar. And half of Terzan 5's certified pulsars are binaries.

图 1 截止 2020 年 12 月双星系统中脉冲星周期的分布统计

Figure 1 Distribution statistics of pulsar periods in binary system till December 2020



截止 2020 年 12 月，ATNF 数据库记录的 149 颗双星系统脉冲星中，有 113 颗属于毫秒级脉冲星，占双星系统的四分之三，约 75.8%。其中，2～4ms 约占毫秒脉冲星总数的六成。

There are 149 binary system pulsars recorded in the ATNF database till December 2020. About 113 are millisecond pulsars, accounting for three-quarters of the binary system (75.8% approximately). It makes up 60% where period is 2～4ms among millisecond pulsars.

根据图 1 的统计结果来看，双星系统中毫秒级脉冲星数量居多。在文[10]中首次提出毫秒脉冲星是正常脉冲星通过在低质量 X 射线双星阶段吸积物质而被加速到毫秒量级。同期发表的文[11]中，Alpar 也认为当低磁场的中子星吸积物质加速到毫秒级后会演化为双星系统。刘鹏等在文[12]中模拟的$P-\dot{P}$图展示了双星脉冲星群自转周期峰值在 5ms 左右，并将双星脉冲星群称为毫秒脉冲星群。由此可知，双星系统与毫秒脉冲星存在密切关系。但数据中心目前针对双星系统的搜索较为耗时，因此，脉冲双星搜索能力的提升十分必要。本文首先介绍脉冲双星系统及其搜索方法，然后分析其方法并根据其特点在程序上实现了并行化，数倍地提高了脉冲双星系统的搜索效率。

1、脉冲双星系统与加速度搜索

1.1 双星系统及搜索技术

脉冲双星系统是相互绕行的两颗中子星，其中一颗是脉冲星，另一颗称为伴星[13]。通常情况下，当星体自转且磁极波束扫过信号探测设备时，探测设备就能够接收到一个脉冲信号[14]。传统的周期搜索中，对采样信号进行快速傅立叶变换将时域信号转换到频域，再通过计算频谱功率来寻找周期。在双星系统中，脉冲星受到伴星引力作用，产生具有加速度的轨道运动，受其影响，脉冲星相对信号探测设备的视向速度将发生变化[8]。由于多普勒效应，探测设备接收到的星体自旋频率将发生变化[9]，传统的周期搜索方法不适用，也难以检测到该脉冲星。

加速度搜索方法[5-7]可有效消除这种影响。该方法采用恒定加速度来近似轨道运动的加速度，从而消除观测数据中由于双星的轨道运动导致的脉冲到达时间的变化[15]。在恒定加速度的假设下，脉冲星的自旋频率在观察者参考系中的漂移速率$\dot{f}$与加速度 α 有如下关系：

$$\dot{f} = \frac{f_0 a}{c} \qquad (1)$$

其中，c 为光速，f_0为脉冲星在自身惯性系中的自旋频率。根据信号处理角度的不同，可分为时域加速度搜索 TDAS（Time Domain Acceleration Search）和频域加速度搜索 FDAS（Fourier Domain Acceleration Search）[16]。下面分别介绍两种方法。

1.2 时域加速度搜索

时域加速度搜索通过线性展宽或压缩信号的方式对数据进行重新采样，以补偿轨道运动引起的多普勒效应，并在一定范围内不断尝试加速度值，来找到最接近真实值的时间序列，再利用周期搜索技术寻找周期信号。进行重采样时信号的时间间隔由公式描述为

$$\tau(t) = \tau_0[1 + \frac{v(t)}{c}] = \tau_0[1 + \frac{\alpha t}{c}] \qquad (2)$$

其中，v 为视向速度，c 是光速。t_0 是保证正确采样的常数，可由公式（2）计算。

$$\tau_0 = \frac{t_{samp}}{1+\frac{aT}{2c}} \quad (3)$$

式中t_{samp}是采样时间间隔，T 是积分时间，α 是尝试的加速度值，c 为光速。

脉冲星搜索软件包 SIGPROC[17]、PEASOUP③等应用了该技术。利用该方法，Amico 等人在球状星团中找到了四个毫秒脉冲星[18]，观测参数见表 2。

表 2 Amico 等人发现的四颗毫秒脉冲星观测参数

Table 2 Observed parameters of four millisecond pulsars discovered by Amico et al.

Pulsar	PSR J1701-30	PSR J1740-53	PSR J1807-24	PSR J1910-59
Period(ms)	5.2415	3.6503	3.0594	3.2661
DM(cm^{-3}pc)	114.4	71.8	134.0	34.0
Orbital period (days)	3.805	1.354	0.071	0.865

1.3 频域加速度搜索

频域加速度搜索基于相关性技术[19]，该技术可提高频谱功率。前文讲到，受轨道运动影响，接收到的脉冲星信号频率将发生变化。这在频谱中表现为某一频率的功率扩散至周围相邻的频率点中。这导致普通的周期搜索方法失效。文[19]中提到，可将原信号与邻近的信号进行相关操作，从而确定相似程度以恢复信号。该过程可由式（4）表示，

$$A_{r_0} \simeq \sum_{k=[r_0]-m/2}^{[r_0]+m/2} A_k A^*_{r_0-k} \quad (4)$$

其中，r_0是要搜索的脉冲星的信号频率点（Fourier frequency bins，可理解为傅立叶域中的频率范围，我们用 0，1，2，3…来代表频谱中的第i个频率点）。A_{r_0}是对应第r_0个频率点的傅立叶响应，其值可由邻近的m个频率点的傅立叶响应来相关地恢复。$A^*_{r_0-k}$是复共轭形式。

通常认为，在观测时间远小于轨道周期的情况下使用该方法效果更好。即满足

③ https://zenodo.org/record/10178

$$T_{obs} < \frac{P_{orb}}{10} \tag{5}$$

其中，T_{obs} 为脉冲星观测时间，P_{orb} 为脉冲星轨道运动周期。当观测时间和轨道周期满足（5）时，加速度可视为常数，结合多普勒公式有：

$$\alpha = \frac{\dot{f}}{f_0}c = \frac{zc}{f_0 T^2} \tag{6}$$

式中，c 是光速，T 为脉冲星观测时间，f 为谐波频率，z 为假设频率漂移的频率点数量。这样，根据漂移的频率点数 z，通过局部信号傅立叶响应的互相关来恢复原始信号。

其他周期搜索方法也可处理该影响，但应用条件不同。相位调制搜索主要针对观测时间与双星轨道周期大致相同的情况，边带搜索在观测时间远大于轨道周期时效果更好。而加速度搜索的使用条件决定了该方法适用目标为周期更小的毫秒脉冲星，在实际中也搜寻到许多优质脉冲星。Ransom 首次使用该方法在球状星团 NGC 6544 中发现脉冲星 PSR J1807−2459A[8]。Barr E.D.利用该技术发现了高度偏心的毫秒脉冲星 PSR J1946+3417，其偏心率达到了 0.13-0.14，这极为罕见同时也预示着不同寻常的形成过程。

2、频域加速度搜索并行化的必要性及可行性分析

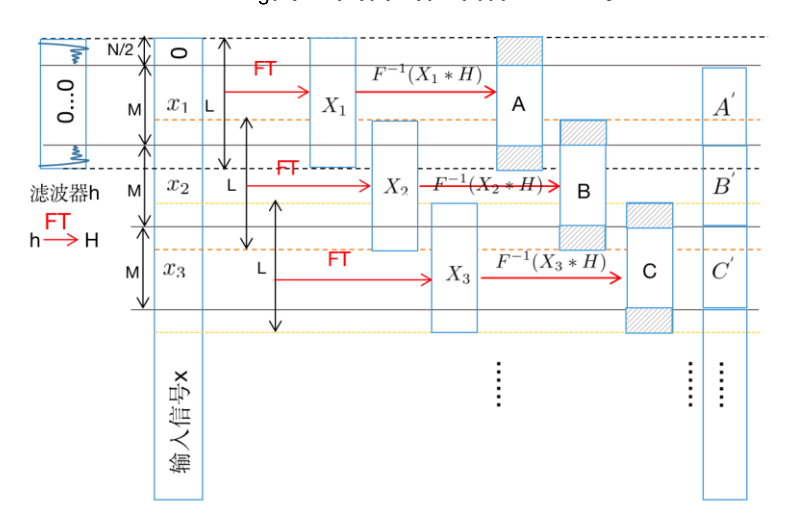
针对频域加速度搜索的并行化提速十分必要。文[15]中论述了轨道运动对脉冲到达时间影响很小，在一定条件下不用加速度搜索也可发现双星。但 2020 年 2 月 FAST 首颗双星的发现中，王琳博士等将 zmax 参数设置为 300 才搜索到脉冲星 PSR J1641+3627F[3]。可见，随着 FAST 搜索要求的不断提升，扩大加速度的搜索范围是必然，否则 FAST 数据无法得到有效吸收，可能错失优质的脉冲星。

相比 TDAS，FDAS 更快且更适合并行化。TDAS 要将处理过的时间序列全部加载到内存，而 FDAS 处理的是局部序列的卷积，更适合小内存的 GPU 进行并行化处理。对尝试的不同加速度值进行处理时，TDAS 需要对每个加速度值重采样后的序列进行 FFT 变换，而 FDAS 只需做一次 FFT 变换，再进行相关操作即可。同时，FDAS 可看作是一个滤波过程，本质是原信号与滤波器的卷积操作。原信号长度远大于滤波器长度，计算的时间复杂度约为 $O(N^2)$，N 为信号长度。为提高计算效率，本文使用快速傅立叶变换做循环卷积，这基于分

治的思想。首先将原信号拆分成与滤波器等长度的信号（滤波器中需要补 0），再分别与滤波器做卷积并做傅立叶逆变换到时域，最后将结果中重叠的部分截去，再合并为输出信号，这一过程也称重叠保留法，可由图 2 描述。这可将算法的时间复杂度降至 O(NlgN)。综上所述，在该过程中使用循环卷积分段计算所具备的局部内存性是高度可并行的，非常适合使用做并行化计算。

图 2 FDAS 中应用的循环卷积④

Figure 2 circular convolution in FDAS



如图，滤波器 $\boldsymbol{h}$ 中间补一定长度的 0 元素，$\boldsymbol{H}$ 为其傅立叶变换。将输入信号 $\boldsymbol{x}$ 分段，每段长为 $\boldsymbol{M}$，将每段信号 $\boldsymbol{x_i}$ 向前多取 $N/2$ 个点，则第一段前面需补 $N/2$ 个 0，再对每段信号做傅立叶变换得到 $\boldsymbol{X_i}$，$\boldsymbol{X_i}$ 与 $\boldsymbol{H}$ 在复数域中相乘再进行傅立叶逆变换得到 $\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$，将 $\boldsymbol{A}$、$\boldsymbol{B}$、$\boldsymbol{C}$ 中混叠的部分截去，得到 $\boldsymbol{A'}$、$\boldsymbol{B'}$、$\boldsymbol{C'}$拼接起来即为等价的线性卷积。

In the figure, the middle of the filter $\boldsymbol{h}$ needs to be filled with a certain length of 0, and $\boldsymbol{H}$ is the Fourier transform. The input signal $\boldsymbol{x}$ is divided into segments, the length of each segment is $\boldsymbol{M}$. Each segment of signal $\boldsymbol{x_i}$ is forwarded by $N/2$ points, thus the first segment needs to be filled with $N/2$ 0 points. Then every signal is transformed to $\boldsymbol{X_i}$ in Fourier domain . $\boldsymbol{X_i}$ and $\boldsymbol{H}$ are multiplied in the complex domain and then inverse Fourier transform is performed to obtain $\boldsymbol{A}$, $\boldsymbol{B}$ *and* $\boldsymbol{C}$, and the aliased part of A, B and C is cut off, and $\boldsymbol{A'}$, $\boldsymbol{B'}$ and $\boldsymbol{C'}$ are spliced together. Lin

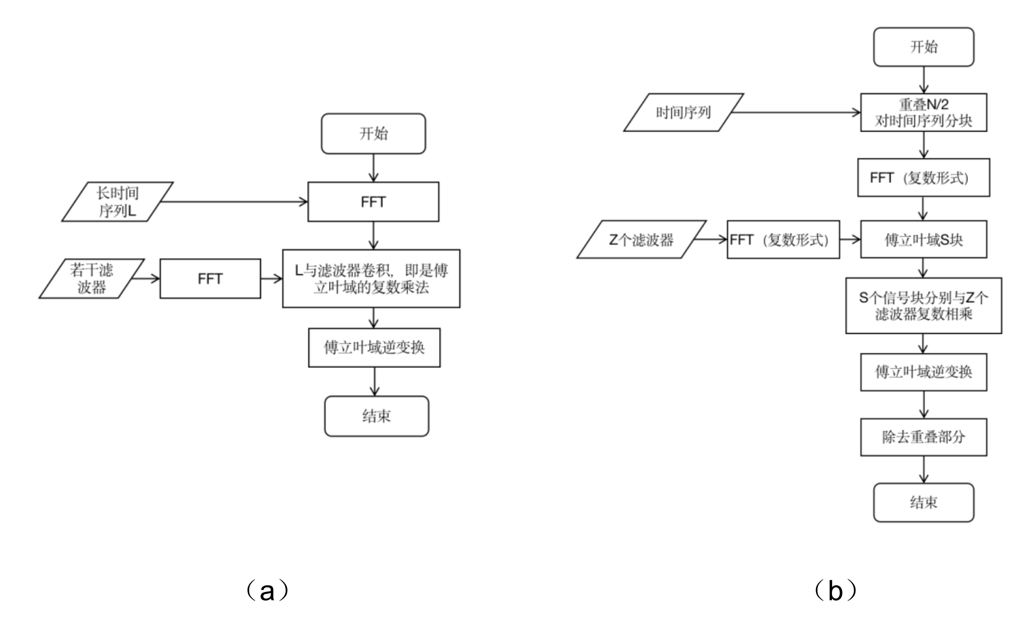
④ https://blog.csdn.net/u010592995/article/details/77882183

ear convolution of valence.

3、频域加速度搜索并行化的实现

根据第二部分的介绍，程序耗时的矛盾主要在于大量可并行的卷积计算。罗晋涛 GPU 版本的加速度搜索在内存方面尚可继续优化。一方面，我们通过循环卷积来局部地缩短时间序列的长度从而降低内存压力。另一方面，循环卷积中分段后的序列可利用 GPU 计算单元的局部内存来完成运算。本文实现的程序流程图见图 3（b）。

在对循环分段的时间序列进行傅立叶变换时，本文使用 CUDA 库封装好的 cuFFT 接口，因为该接口已针对 NVIDIA GPU 进行了 FFT 的高度优化。利用批处理 cuFFT 例程，在 GPU 上并行调用一次即完成多段序列的傅立叶变换。处理加速度滤波器时，我们在一个 GPU Grid 中加载滤波器，其中每一个线程中都加载全部滤波器，然后沿分段处理并经过傅立叶变换后的信号，分别与不同的分段频域信号进行复数域的乘法运算。这一过程在局部进行，且每个线程都在同步处理相同的运算。再将运算结果写回到全局内存中进行傅立叶逆变换和重叠段的去除，进行合并后得到输出。



（a）　　　　　　　　　　　　　　　　（b）

图 3. （a）线性卷积运算流程;（b）本文中循环卷积处理流程

Fig 3. (a) The flow chart of linear convolution; (b) The process of cyclic convolution in this article.

4、实验及结果

实验使用 AMD Ryzen7 2700X 8 核 16 线程 CPU，两张 GeForce GTX1080 显卡，对 FAST 漂移扫描数据的若干文件（PSR20180811_00A01H、PSR20180826_00A01H、PSR2018

0828_00A01H、PSR20180902_00A01H，观测参数见表 3）消色散后产生的.dat 文件进行测试，使用不同 zmax 参数对该批文件处理，产生相应标记周期信号的 ACCEL 文本文件，并与 CPU 版本、罗晋涛 GPU 版本的加速度搜索做了对比。

表 3. 本次实验使用的观测文件参数

Table 3. Observed parameters used in this experiment

Parameter	Value
Telescope used	FAST
Instrument used	MB4K
Epoch of observation (MJD)	58341.09509（PSR20180811_00A01H）
	58356.06175（PSR20180826_00A01H）
	58358.07425（PSR20180828_00A01H）
	58363.02564（PSR20180902_00A01H）
Number of bins in the time series	70902784
Width of each time series bin (sec)	4.9152e-05
Central freq of low channel (MHz)	1124.93896484
Total bandwidth (MHz)	250
Number of channels	2048
Channel bandwidth (MHz)	0.1220703125

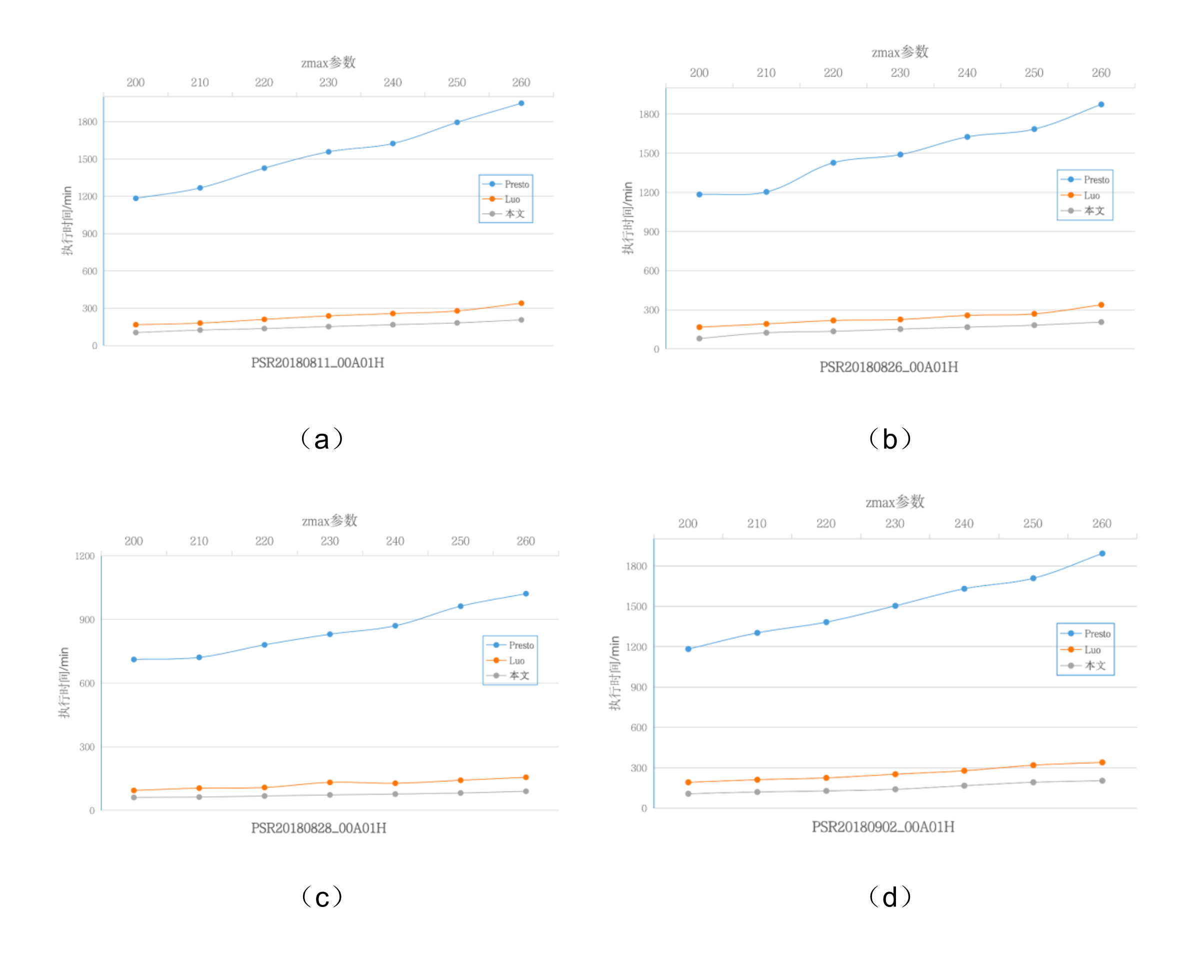


图 4. Ransom 版本、罗晋涛版本与本文方法使用不同 zmax 参数对四个文件的执行所消耗的时间对比结果。其中， (a)、(b)、(c)、(d)依次对应四个文件 PSR20180811_00A01H、PSR20180826_00A01H、PSR20180828_00A01H、PSR20180902_00A01H。

Fig4. Comparison of different methods in execution time. (a)PSR20180811_00A01H;(b)PSR20180826_00A01H; (c)PSR20180828_00A01H; (d)PSR20180902_00A01H.

结果表明，采用文中方法对加速度搜索程序改进后，本文并行化处理方法耗时相比 Ransom 版本有极大提升，加速比约 10-12 倍，较罗晋涛版本也有一定进步，提升约 1.5-1.7 倍。综合来看，程序的改进有一定成效，达到预期效果。

致谢：特别感谢实验室刘志杰老师和于徐红老师在程序修改上的帮助。

参考文献：

[1] Hulse R A, Taylor J H. Discovery of a pulsar in a binary system [J]. The Astrophysical Journal, 1975, 195: L51-L53.

[2] J.H.Taylor,L.A.Fowler,P.H.McCulloch,高学贤,戴经维.测量脉冲双星 PSR1913+16 的广义相对论效应[J].世界科学译刊,1979(07):21-26.

[3] Wang L, Peng B, Stappers B W, et al. Discovery and Timing of Pulsars in the Globular Cluster M13 with FAST[J]. The Astrophysical Journal,2020,892(1).

[4] 张辉, 谢晓尧, 李药, 等. 一种面向 FAST PB 量级脉冲星数据处理加速方法及系统[J]. 天文研究与技术, 2020，18(1): 129-137.

[5] Pan Z.,Hobbs G.,Li D.,Ridolfi A.,Wang P.,Freire P.. Discovery of two new pulsars in 47 Tucanae (NGC 104)[J]. Monthly Notices of the Royal Astronomical Society: Letters,2016,459(1).

[6] Zhichen Pan,Scott M. Ransom,Duncan R. Lorimer,William C. Fiore,Lei Qian,Lin Wang,Benjamin W. Stappers,George Hobbs,Weiwei Zhu,Youling Yue,Pei Wang,Jiguang Lu,Kuo Liu,Bo Peng,Lei Zhang,Di Li. The FAST Discovery of an Eclipsing Binary Millisecond Pulsar in the Globular Cluster M92 (NGC 6341)[J]. The Astrophysical Journal Letters,2020,892(1).

[7] Andersen B C, Ransom S M. A Fourier Domain “Jerk” Search for Binary Pulsars[J]. The Astrophysical Journal Letters, 2018, 863(1): L13.

[8] Ransom S M, Greenhill L J, Herrnstein J R, et al. A binary millisecond pulsar in globular cluster NGC 6544[J]. The Astrophysical Journal Letters, 2000, 546(1): L25.

[9] Lorimer D R. Binary and millisecond pulsars at the new millennium[J]. Living reviews in relativity, 2001, 4(1): 1-87.

[10] Backer D C, Kulkarni S R, Heiles C, et al. A millisecond pulsar[J]. Nature, 19

82, 300(5893): 615-618.

[11] Alpar M A, Cheng A F, Ruderman M A, et al. A new class of radio pulsars[J]. Nature, 1982, 300(5894): 728-730.

[12] 刘鹏,王培,李菂,张洁,张蕾,张承民,朱炜炜,岳友岭,代实.FAST 19 波束脉冲星漂移扫描巡天模拟[J].天文学进展,2018,36(02):173-188.

[13] 李向东,汪珍如.X 射线脉冲双星的观测特性[J].天文学进展,1993(02):106-116.

[14] 帅平，陈绍龙，吴一帆，等. X 射线脉冲导航原理[J]. 宇航学报，2007，28(6)：1538.

[15] 潘之辰,钱磊,岳友岭.脉冲星搜索技术及 FAST 望远镜脉冲星搜索展望[J].天文研究与技术,2017,14(01):8-16.

[16] Cameron A D. Innovative Pulsar Searching Techniques[J].威廉姆斯大学，2018.

[17] Lorimer D R. SIGPROC: pulsar signal processing programs[J]. Astrophysics Source Code Library, 2011, 1107.016.

[18] D'Amico N, Lyne A G, Manchester R N, et al. Discovery of short-period binary millisecond pulsars in four globular clusters[J]. The Astrophysical Journal Letters, 2001, 548(2): L171.

[19] Ransom S M, Eikenberry S S, Middleditch J. Fourier techniques for very long astrophysical time-series analysis[J]. The Astronomical Journal, 2002, 124(3): 1788.

Binary Pulsar System Acceleration Search Method and Software Improvement

Ji Changxu，Liu Zhijie，Yu Xuhong

Key Laboratory of Information and Computing Science in Guizhou Province

Abstract: Millisecond pulsars are of great significance to gravitational wave detectio n and space navigation. We point out that the binary star system is closely related to the millisecond pulsar by citing the ATNF database. We analyzed the search c apabilities of binary pulsar system in early scientific data center. Although the curre nt version of the accelsearch is highly optimized on CPU, it still consumes computi ng power under the huge amount of FAST data. Luo's GPU version of the acceler ation search brings 6-8 times speedup, but it fails to optimize the short signal and GPU memory effectively. We summarize the impact of orbital motion in binary puls ar system firstly, and then introduce the search technology. We write programs to optimize the GPU version of the acceleration search in order to improve the search efficiency. Our benchmark shows that using different zmax parameters to process different .dat files has significantly improved the speed.

Key words：Acceleration Search; Binary Pulsar System; FAST